TD4 de Statistique Inférentielle

Test de Neyman-Pearson et Applications

Magistère d'économiste statisticien

1 Loi de Pareto

Soit $X_1, \dots, X_n$ un n-échantillon de densité :

$$f(x) = \theta x^{-(1+\frac{1}{\theta})} \mathbf{1}_{[1,+\infty[}(x), \tag{1}$$

où $\theta > 0$ est un paramètre inconnu. On pose pour $j = 1, \dots, n$, $Y_j = \log X_j$ et $\overline{Y}_n = \frac{1}{n} \sum_{j=1}^{n} Y_j$.

1) Montrer que Y_1 est de carré intégrable. Calculer l'espérance et la variance de Y_1. En déduire que $\overline{Y}_n$ converge, quand n tend vers l'infini, presque sûrement vers θ, et que $\sqrt{n}(\overline{Y}_n - \theta)$ converge en loi vers une loi que l'on précisera.
2) Calculer la fonction log-vraisemblance associée aux observations $X_1, \dots, X_n$ et l'estimateur du maximum de vraisemblance $\hat{\theta}_n$. Cet estimateur est-il sans biais ? Quelle est sa variance ? Montrer qu'il converge presque sûrement vers θ.
3) Pour n grand, donner un intervalle de confiance de risque α ($\alpha \in]0, 1[$). Application numérique : $\alpha = 5\%$, $n = 40$, $\prod_{j=1}^{n} x_j = 1000, 5$.
4) On veut tester $H_0 \quad \theta = 2$ contre $H_1 \quad \theta = \frac{5}{2}$. Montrer que la stratégie optimale consiste à décider H_1 si $\overline{Y}_n > C_n$ et H_0 sinon. On suppose que $n = 50$ et on fixe l'erreur de première espèce à 5%. Calculer :
   - Une approximation de C_n.
   - Une approximation de l'erreur de seconde espèce.

2 Vraisemblance gaussienne

Soit $X_1, \dots, X_n$ un n-échantillon de la loi $\mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$. On note Φ la fonction de répartition de la loi normale centrée réduite. Soit $\alpha \in]0, 1[$ et t_α défini par l'équation $\Phi(t_\alpha) = 1 - \alpha$.

a) Ecrire le test du rapport de vraisemblance au niveau α de "$\theta = 0$" contre "$\theta > 0$". Calculer la fonction de puissance de ce test.
b) Déterminer la zone de rejet du test du signe au niveau α de "$\theta = 0$" contre "$\theta > 0$". Exprimer sa fonction de puissance.
c) Trouver un équivalent de la fonction de puissance du test du signe lorsque n tend vers l'infini et θ tend vers zéro. (Cet équivalent s'exprime à l'aide de Φ, t_α, n et θ).
d) On pose $\alpha = 0, 05$ et $\theta_1 = 0, 1$.
   α) Soit $n = 100$. Calculer la puissance du test de rapport de vraisemblance au point $\theta = \theta_1$.
   β) En utilisant l'approximation établie au c), évaluer le nombre minimum d'observations nécessaires pour garantir une puissance identique à celle calculée en α) mais cette fois pour le test du signe (toujours au point $\theta = \theta_1$). Qu'en pensez-vous ?

3 Lois hypergéométriques

Pour $N \in \mathbb{N}_*$ et $n \leq N$ fixés, on considère $\Theta = \{0, 1, \dots, N\}$ et la famille des lois hypergéométriques sur $\Omega = \{0, 1, \dots n\}$, $(P_\theta)_{\theta \in \Theta}$,

$$P_\theta(k) = \frac{\begin{pmatrix} \theta \\ k \end{pmatrix} \begin{pmatrix} N - \theta \\ n - k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}, \qquad \max(0, n - (N - \theta)) \leq k \leq \min(n, \theta).$$

On observe X de loi P_θ. Bâtir un test uniformément le plus puissant de $\theta \leq \theta_o$ contre $\theta > \theta_o$.

4 Neyman-Pearson

Soient X une variable aléatoire réelle de loi $\mathcal{B}(n, p)$ et φ un test de Neyman-Pearson $\varphi = \mathbb{1}_{X > c} + \gamma \mathbb{1}_{X = c}$ de $p \leq p_o$ contre $p > p_o$, de taille α.

a) Pour $n = 6, p_o = 0,25$ et les niveaux $\alpha = 0,05; 0,1\ 0,2$, déterminer c, γ et calculer la puissance du test contre $p_1 = 0,3\ 0,4\ 0,5\ 0,6\ 0,7$.

b) Soient $\alpha = 0,05$ et $p_o = 0,2$. Si on désire une puissance $\beta \leq 0,9$ contre $p_1 = 0,4$, calculer le nombre n d'observations nécessaires en utilisant tantôt des tables de la loi binômiale, tantôt l'approximation gaussienne.

c) Utiliser l'approximation gaussienne pour calculer le nombre d'observations requises pour avoir $\beta \geq 0,9$ quand $\alpha = 0,05$, $p_o = 0,01$ et $p_1 = 0,02$.

5 Rapports de vraisemblance monotones

Soit f une densité strictement positive sur $\mathbb{R}$. Montrer qu'une condition nécessaire et suffisante pour que la famille $\{f(\cdot - \theta)\}_{\theta \in \mathbb{R}}$ soit à rapport de vraisemblance croissant est que $\log(f)$ soit concave.

6 Lois uniformes

Soit $X_1, \dots, X_n$ un n-échantillon de la loi uniforme sur $[0, \theta]$. Montrer que

a) pour tester $\theta \leq \theta_o$ contre $\theta > \theta_o$ au niveau α, tout test φ de niveau α tel que $E_{\theta_o}\varphi(X) = \alpha$ et $\varphi(x) = 1$ lorsque $\max(x_1, \dots, x_n)$ est un test uniformément le plus puissant ,

b) il existe un unique test uniformément le plus puissant de $\theta = \theta_o$ contre $\theta \neq \theta_o$ au niveau α, donné par $\varphi(x) = 1$ lorsque $\max(x_1, \dots, x_n) > \theta_o$ ou $\max(x_1, \dots, x_n) \leq \theta_o \alpha^{1/n}$ et $\varphi(x) = 0$ sinon.

7 Exponentielles décentrées

Soit $X_1, \dots, X_n$ un n-échantillon de la loi de densité $x \mapsto a \exp(-a(x - b))\, \mathbb{1}_{(x \geq b)}$.

a) On suppose que a est connu. En utilisant l'exercice précédent, montrer qu'il existe un unique test uniformément le plus puissant de $b = b_o$ contre $b \neq b_o$.

b) Montrer qu'il existe un unique test uniformément le plus puissant de $a = a_o$, $b = b_o$ contre $a > a_o$, $b < b_o$. Expliquer l'existence (inhabituelle pour un test à deux paramètres) de ce test uniformément le plus puissant .